Technical Report 1004

Creation of New Items and Forms for the Project A Assembling Objects Test

Henry H. Busciglio, Dale R. Palmer, Ivey H. King, and Clinton B. Walker
U.S. Army Research Instutute

August 1994

United States Army Research Institute for the Behavioral and Social Sciences

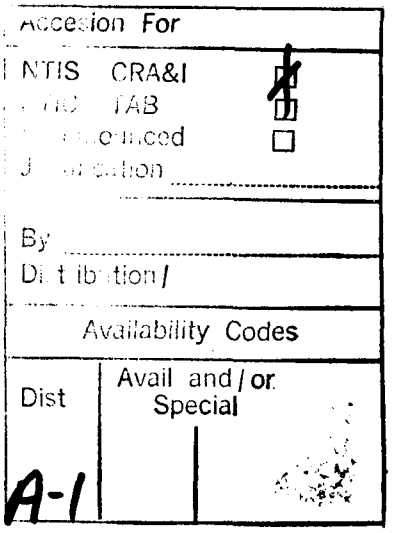
U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Technical review by

Elizabeth J. Brady
Judith E. Brooks



NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1994, August	Final Oct 90 - Jan 94

4. TITLE AND SUBTITLE
Creation of New Items and Forms for the Project A Assembling Objects Test

5. FUNDING NUMBERS
PE 63007A
792
1222H
879300

6. AUTHOR(S)
Busciglio, Henry H.; Palmer, Dale R.; King, Ivey H.; and Walker, Clinton B.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-RS
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

8. PERFORMING ORGANIZATION REPORT NUMBER
ARI Technical Report 1004

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
--

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
--

11. SUPPLEMENTARY NOTES
--

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT *(Maximum 200 words)*

The Army's Project A was a comprehensive effort to improve the selection and classification of enlisted personnel. The Assembling Objects (AO) test was a major product of this effort. Previous research has shown AO to be an excellent measure of both overall spatial ability and complex, g-loaded problem-solving skills. In view of the great potential usefulness of the AO measure, researchers from the U.S. Army Research Institute for the Behavioral and Social Sciences examined the original test to develop precise, comprehensive item specifications. Using these specifications, they developed new draft items that were psychometrically tested in a field setting. The researchers then chose the best of the new items and combined them into three complete new forms that were further tested. Analyses showed that all three new forms displayed acceptable (or better) psychometric properties, at both the item and total score level, thus supporting the usefulness of the item specifications for creating new AO items and forms. These efforts should help to make the Assembling Objects test a valuable addition to the testing programs of the Army and the other armed services.

14. SUBJECT TERMS
Project A
Spatial aptitude
Spatial visualization

15. NUMBER OF PAGES
48

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500
Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18
298-102

Technical Report 1004

Creation of New Items and Forms for the Project A Assembling Objects Test

Henry H. Busciglio, Dale R. Palmer, Ivey H. King, and Clinton B. Walker
U.S. Army Research Instutute

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

Manpower and Personnel Research Division
Zita M. Simutis, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

August 1994

Army Project Number
2Q263007A792

Manpower and Personnel

FOREWORD

The Army's Project A was a comprehensive effort to improve the selection and classification of enlisted personnel. The Assembling Objects test was a major product of this effort. This instrument has been shown to be an excellent measure of both overall spatial ability and complex, g-loaded problem-solving skills. A number of analyses have demonstrated its superior performance in incrementing the validity of the Armed Services Vocational Aptitude Battery (ASVAB) across a wide range of criteria and Military Occupational Specialties.

Because of its very good psychometric properties, the Assembling Objects test has been included in the joint-services Enhanced Computer Administered Testing project as a candidate for inclusion in future versions of the ASVAB. In view of the great potential usefulness of the Assembling Objects measure, U.S. Army Research Institute for the Behavioral and Social Sciences researchers

- examined the original test to develop precise, comprehensive item specifications;
- used these specifications to develop new draft items that were psychometrically tested in a field setting;
- chose the best of the new items and combined them into three complete new forms that were further tested in a field setting.

These efforts should help to make the Assembling Objects test a valuable addition to the testing programs of the Army and the other armed services.

EDGAR M. JOHNSON
Director

ACKNOWLEDGMENTS

We would like to express our appreciation to the following individuals for their valuable assistance:

- Personnel of the 43rd AG Battalion at Fort Leonard Wood and the Training, Development, and Analysis Directorate of Training and Doctrine Command for arranging the data collections.

- The American Institutes for Research and RGI, Inc., for creating test materials, collecting the psychometric data, and creating the data files.

Special thanks also to Michael Rumsey, Selection and Assignment Research Unit Chief, and to Elizabeth Brady and Judith Brooks for their astute and constructive comments on earlier drafts of this paper.

CREATION OF NEW ITEMS AND FORMS FOR THE PROJECT A ASSEMBLING OBJECTS TEST

EXECUTIVE SUMMARY

Requirement:

This research developed specifications for creating new items and forms for the Army's Assembling Objects test and created these items and forms and evaluated their psychometric properties. The Assembling Objects test has been shown to be an effective predictor of job performance, over and above the Armed Services Vocational Aptitude Battery (ASVAB; e.g., Busciglio, 1990). The test has been included in the joint-services Enhanced Computer Administered Testing (ECAT) project to be evaluated as a potential subtest in future versions of ASVAB. It is therefore desirable to create an expanded pool of items and/or one or more entire new forms of this measure.

Procedure:

Researchers from the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) examined the 36 items on the original form of the Assembling Objects test to develop a set of item specifications to guide the creation of 144 new draft items. After numerous quality control reviews and refinements, these items were administered to a sample of new Army recruits, along with the 36 original items. Based on item analyses, we retained 108 of the new items and combined them into three new alternate forms of the test. In a second data collection, each of the three proposed alternate forms was administered with the original form.

Findings:

Analyses showed that all three new forms displayed acceptable (or better) psychometric properties, at both the item and total score level, thus supporting the usefulness of the item specifications for creating new Assembling Objects (AO) items and forms.

Utilization of Findings:

If the Assembling Objects test becomes part of the Army's selection and classification procedures, the creation of psychometrically sound alternate forms will ensure the long-term validity and usefulness of this measure. The test specifications, the original form, and the new items and forms of Assembling Objects are a sound foundation for bringing this measure into future versions of ASVAB.

CREATION OF NEW ITEMS AND FORMS FOR THE PROJECT A ASSEMBLING OBJECTS TEST

CONTENTS

CONTENTS (Continued)

Page

LIST OF TABLES

CONTENTS (Continued)

LIST OF FIGURES

Creation of New Items and Forms for the Project A Assembling Objects Test

Introduction

Development of the Original Assembling Objects Test

Project A (e.g., Campbell, 1989) was a long-term, comprehensive program of research and development aimed at improving the Army's selection and classification systems through the creation and validation of measures of skills and abilities other than those tapped by the Armed Services Vocational Aptitude Battery (ASVAB). As a result of an exhaustive search undertaken during the early stages of Project A, researchers determined that spatial aptitude was one of the most promising of these skills and abilities. The original form of the Assembling Objects test was among the instruments created to tap this dimension. Specifically, it was designed to measure a construct called 'Spatial Visualization/Rotation,' defined as

> the ability to identify a two-dimensional figure when seen at different angular orientations within the picture plane. It also includes three-dimensional rotation or the ability to identify a three-dimensional object projected on a two-dimensional plane, when seen at different angular orientations either within the picture plane or about the axis in depth (Peterson, 1987, p. 3-5).

The contractors responsible for the development of the original Assembling Objects (AO) test identified a number of published "marker" tests of this ability, such as the Assembly subtest of the Flanagan Industrial Test (FIT; Flanagan, 1965) and the Space Visualization subtest of the Employee Aptitude Survey (EAS-5; Ruch & Ruch, 1980). As explained by Peterson (1987), marker tests were

> judged to measure the predictor categories on constructs for which we were developing tests. Some of these marker tests were actually administered during pilot testing, others were not, but they were all studied to assist in developing the new tests (p. 3-6).

As part of Project A, expert raters determined that a measure similar to the marker tests would have predictive validity for MOS that involve mechanical operations, construction, and drawing/using maps (Peterson, 1987).[1]

---

[1]We might add that one type of item on the Assembling Objects test (the "puzzle" items, as described below) are also similar to a very popular occupational measure of spatial aptitude, the Minnesota Paper Form Board Test (MPFB; Likert & Quasha, 1970; also see Anastasi, 1982). Alderton (1989b, p. 5) has reported that

During the course of test development, a number of different versions of the AO measure were tried out (Peterson, 1987; Figure 1 shows a brief summary of the development of the original Assembling Objects test). The final version consists of 36 items and has an 18 minute time limit. The score is simply the number of items answered correctly, with no penalty for guessing.[2] To help readers understand the exact nature of the test, the instructions, including a number of example items, are shown in Figure 2.

Psychometric Properties of the Original Assembling Objects Test

This section is concerned with empirical data on the potential usefulness of including the Assembling Objects test in future versions of the ASVAB. It will cover three basic topics:

1) Evidence for the psychometric uniqueness of AO from ASVAB.

2) The incremental validity of AO scores relative to the ASVAB.

3) Correlational and factor analytic evidence of what AO measures; particularly, evidence that AO is:

   a) an excellent 'summary measure' of spatial ability that can be quite useful in situations in which limited resources preclude the use of more than one measure;

   b) a complex reasoning test that is related to general cognitive ability (g) and can therefore be expected to predict many aspects of job performance, including nonspatial components.

Psychometric Uniqueness of AO. For a measure to be a useful addition to ASVAB, it must tap abilities other than those already being measured by operational Army tests. That is, a nontrivial part of its reliable variance must be uncorrelated with the ASVAB measures.

In his report on the development of the Project A measures, Peterson (1987) reported data gathered on an early (40-item) version of the AO test. The zero-order correlations found between AO and the various ASVAB subtests were low to moderate, ranging from -.01 (Number Operations) to .57 (Mechanical

---

complex power tests of spatial ability such as the MPFB "are significantly related to standard academic course grades, vocational course work, and job success across the entire occupational spectrum."

[2]During the development of the AO test, a number of test parameters (e.g., number of items, time limits) and scoring procedures were explored; see Peterson (1987) and Campbell and Zook (1992) for more complete discussions of the decision process that led to the final version of the test.

Data From:	# Items	Time limit (Min.)	Sample size	Results and changes made
First Tryout	30	16	36	Ceiling effects noted/ ten harder items added Correlations with marker tests: FIT - .76 EAS - .74
Second Tryout	40	16	56	Ceiling effects reduced Correlation with marker test: FIT - .64
Pilot Test	40	16	118	Some items redrawn to clarify figures Item response format changed for machine scoring
Field Test	40	16	168	Eight items dropped, based on poor item characteristics Reliability estimates: alpha - .92 split half - .79 test-retest - .74
Project A Concurrent Validation	32	16	9,345	Ceiling effects noted/ four items added, three revised Time limit increased by two min. Reliability estimates: alpha - .90 test-retest - .70
Career Force Longitudinal Validation	36	18	6,950	Ceiling effects reduced Reliability estimate: alpha - .88

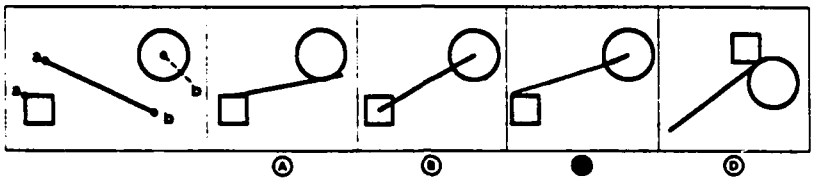
Figure 1. A brief summary of the development of the original Assembling Objects test.

In this test you are to figure out how an object will look when its parts are put together correctly.

For each problem, the first picture shows the parts that need to be put together. The next four pictures, labeled A, B, C and D, show four different ways the parts could be put together. Only one of them is correct.

There are two types of problems in the test. In the first type, the parts are labeled with letters, and by matching the letters, you can see where the parts should touch when the object is put together correctly. Look at Example 1.

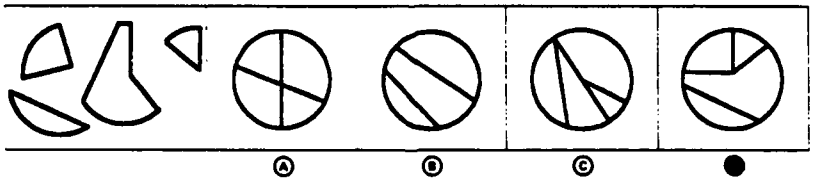
Example 1:



First, look at the square with the corner marked "a." Imagine what the object would look like if you attached that corner to the end of the line marked "a," and then attached the "b" end of the line to the middle of the circle, where the "b" is. Of the four answer choices, labeled A, B, C, and D, only C shows the object put together correctly, so circle C has been filled in.

For the second type of problem in this test, the parts to be put together are not labeled. In these problems, you just fit the parts together like the pieces of a puzzle. Only one of the four answer choices is correct. Look at Example 2 below.

Example 2:



When you piece together the parts of the first picture, only answer D can be correct. D is the only answer choice that contains all the right shapes or pieces of the puzzle.

To mark your answer, fill in the lettered circle below the correct answer choice, as shown in the two examples above.

Figure 2. Instructions for the original Assembling Objects test.

Comprehension), with a median of .39. The correlation with overall AFQT score was .44.[3]

Peterson (1987) also reported a "uniqueness estimate" for the AO test of .39; this was defined as the proportion of total variance that was reliable but not related to the ASVAB (more technically, the split-half reliability minus the R^2 with the ASVAB subtests in a regression analysis). In his report on the Project A Concurrent Validation, during which a 32-item version of the AO test was used, Campbell (1988) noted a uniqueness estimate of .65. The measures developed in the Army's Project A were administered to a very large sample of Army recruits in 1986-87, as part of another comprehensive research effort, a longitudinal validation study called Building the Career Force. Analyzing predictor data from this effort, Campbell and Zook (1992) reported a uniqueness estimate of .62 [split-half reliability estimate of .90 minus an R^2 with all ASVAB subtests of .28].[4]

Incremental Validity of AO. Psychometric uniqueness is a necessary but not sufficient condition for incremental validity; not only must a test measure abilities not tapped by ASVAB, but these abilities must be significantly related to important aspects of job performance. Analyzing the Project A Concurrent Validity data, Busciglio (1990) found that the AO test was superior to all other Project A spatial and perceptual-psychomotor measures for incrementing ASVAB validity, across a broad range of MOS and performance criteria. Specifically, the author employed stepwise regression analyses to assess a total of 132 predictor-criterion relationships in which the Project A measures could increment the validity of an empirically determined best possible combination of ASVAB subtests. In these procedures, the ASVAB subtests were entered first, significant subtests were retained, and then the Project A tests were entered and significant incremental predictors were determined. The AO test was significant in 61 out of a possible 132 equations, the highest number among all Project A spatial and perceptual-psychomotor tests.[5] Even when compared with ASVAB subtests,

---

[3]The AFQT (Armed Forces Qualification Test) score is a composite of four ASVAB verbal and mathematics subtests. A percentile score - relative to the national youth population - is one of several criteria used to make selection decisions.

[4]The most recent, 36-item version of AO was used.

[5]The next highest number was 52, for another spatial test, Figural Reasoning. The highest number for a perceptual-psychomotor test was 45, for the % correct score on the Target Identification test, a measure of perceptual speed and accuracy.

AO's validity was high; AO was superior to six of the nine ASVAB subtests for predicting the various criteria.[6]

What AO Measures: Spatial Ability. Campbell and Zook (1992) reported a series of factor analyses that included the AO measure. Data were from the Project A Concurrent Validation sample and two different samples from the predictor data collection in the Career Force Longitudinal Validation. The first set of factor analyses included the six Project A spatial tests. As Table 1 shows, results were very similar in all three samples. Although all three analyses showed a two-factor solution, the AO and Figural Reasoning tests cross-loaded on the two factors.[7]

Table 1

Factor Analysis Results for Assembling Objects and Other Project A Spatial Tests (from Campbell & Zook, 1992)

	Loadings					
	Factor I			Factor II		
Test	CV	LV1	LV2	CV	LV1	LV2
Assembling Objects	.54	.55	.54	.47	.49	.50
Figural Reasoning	.59	.54	.54	.40	.46	.42
Map	.60	.59	.58	.37	.38	.38
Orientation	.56	.57	.56	.34	.37	.35
Maze	.38	.38	.37	.57	.57	.55
Object Rotation	.32	.36	.34	.52	.54	.52

Note. CV is Concurrent Validation sample from Project A (N = 7939). LV1 and LV2 are two different samples from the predictor data collection phase of Project A Longitudinal Validation (Ns = 6929 and 6436, respectively).

The same authors reported further analyses on the first Longitudinal Validation sample, using the Schmid-Leiman (1957, as cited in Campbell & Zook, 1992) transformation. Results are shown in Table 2. As can be seen, an oblique rotation yielded a

[6]AO significantly incremented ASVAB validity in 61 equations. Only three ASVAB tests were initially significant in a greater number of equations; these were Mathematics Knowledge (96 equations), Auto/Shop Information (94), and Mechanical Comprehension (70).

[7]The authors asserted that the factor formed by the Maze and Object Rotation tests "appears to be a method (speededness) factor that does not reflect a meaningful homogeneous construct" (p. 94); thus, they concluded that these analyses actually support a one-factor solution.

Table 2

Schmid-Leiman Results for Assembling Objects and Other Project A Spatial Tests (from Campbell & Zook, 1992)

	Loadings			
	General	Specific Factors:		
Test	Factor	Speed	Figural	Orientation
Assembling Objects	.753	.000	.065	.000
Figural Reasoning	.720	.000	.062	.000
Map	.685	.000	.000	.278
Orientation	.656	.000	.000	.266
Maze	.624	.367	.000	.000
Object Rotation	.592	.347	.000	.000

Note. Sample from predictor data collection phase of Project A Longitudinal Validation (N = 4723).

three-factor solution, but all tests loaded more highly on a second-order general factor, with AO having the highest loading.

Campbell and Zook (1992) also reported a number of factor analyses including the six Project A spatial tests and the ASVAB subtests. Analyses of both Concurrent and Longitudinal Validation data showed the six spatial tests loading cleanly on the same factor, with the AO test having the highest factor loading.[8] Based on all their analyses, the authors concluded:

> For applied settings in which the Army may wish to use fewer than six tests to test spatial abilities, we suggest using the Assembling Objects Test. It is a good measure of the general [spatial] factor, and it consistently yields smaller gender and race differences than the other tests (p. 101).

What AO Measures: General Cognitive Ability. Busciglio's (1990) finding that the AO test is excellent for incrementing ASVAB validity, as described above, is also evidence for its usefulness as a measure of general cognitive ability. In another analysis of Project A Concurrent Validation data, Busciglio (1991) found that the AO test was a valid predictor of a variety of different Project A comprehensive performance measures. Table 3 shows a summary of results from both of Busciglio's analyses. As can be seen, AO was superior to seven of the nine ASVAB subtests in predicting performance across criterion measures and

[8]The only ASVAB subtest to load on the spatial factor was Mechanical Comprehensive, which cross-loaded on an ASVAB Technical factor that also included the Auto/Shop and Electronics subtests.

Table 3

Number of Army MOS for Which Assembling Objects and ASVAB Subtests Are Significant Predictors of Comprehensive Performance Measures (from Busciglio, 1990 and Busciglio, 1991)

	Performance Measures				
	Written Knowledge	General Soldiering	Core Technical	Skill Qual	TOTAL
Assembling Objects	7	7	6	3	23
ASVAB:					
Auto/Shop	7	8	7	7	29
Math Knowledge	8	7	6	4	25
Mechanical Comp.	7	5	5	3	20
Verbal	7	2	5	5	19
General Science	4	5	3	6	18
Arithmetic Reas.	2	2	3	5	12
Number Operations	4	2	2	1	9
Electronics Info.	3	0	4	2	9
Coding Speed	2	1	4	0	7

Note. A total of 9 entry-level MOS comprised the sample. Data on Assembling Objects are from Busciglio (1991); results for ASVAB are from Busciglio (1990). Criterion measures were as follows: Written Knowledge - multiple-choice tests of technical knowledge. General Soldiering - written and hands-on tests of tasks common to many MOS. Core Technical - written and hands-on tests of the tasks that define the MOS. Skill Qualification - paper-and-pencil tests of MOS-specific technical knowledge developed by the U.S. Army Training and Doctrine Command for periodic testing of soldiers in their MOS.

Military Occupational Specialties (MOS). It is also worth noting that, in 21 of the 23 regression equations where it was initially significant, the AO test remained significant after the ASVAB subtests were entered. It would appear, then, that the complex, problem-solving nature of the AO test taps general cognitive abilities that are not currently being measured by the ASVAB.[9]

Inclusion of Assembling Objects in the Enhanced Computer Administered Testing Project

Because of results such as those above, the Military Accessions Policy Working Group decided to include the original AO test in an experimental battery of computerized spatial and perceptual-psychomotor measures. The incremental validity of this battery, over and above the ASVAB, is being assessed in a joint services research effort called the Enhanced Computer

[9]Lohman (1979) has provided an exhaustive review and reexamination of factor analytic studies of spatial measures, including evidence for the strong empirical relationship between complex tests of 'Visualization' - such as AO - and more general cognitive ability constructs.

Administered Testing (ECAT) project. Preliminary analyses of ECAT data further support the potential usefulness of the AO test.

For example, Carey (1992) assessed the validity of the ECAT tests for predicting a hands-on measure of mechanical maintenance performance, in a sample of Marine automotive and helicopter mechanics.[10] The author reported that AO was, across a number of analyses, superior to all other ECAT tests as a predictor and incremental predictor (over ASVAB and time-in-service). The AO test was also superior to a number of psychomotor and dexterity subtests of the General Aptitude Test Battery (GATB; e.g., Hartigan & Wigdor, 1989).

Overall, these analyses show that AO is an excellent measure of both overall spatial ability and complex, g-loaded problem-solving skills. These analyses have also demonstrated AO's superior performance in incrementing the validity of the ASVAB across a wide range of criteria and Military Occupational Specialties. In light of this evidence for the potential usefulness of the AO measure, we determined that precise item specifications, as well as additional items and forms of the test should be developed. The remainder of this paper describes the methods we used and the results we obtained.

Creation of Original Pool of Draft Items

Formulation of Item Specifications

As noted above, the AO test consists of two types of items - Items with Marked Connections (see Example 1 in Figure 2) and Puzzle Items (Example 2 in Figure 2). The test was content analyzed to determine the characteristics of the original items. After discussing the importance of these characteristics, we developed item specifications that included the important characteristics and specified the number of each type of item to be developed. Figure 3 shows the characteristics of the original items. Specifications for the new items are shown in Figure 4.[11]

---

[10] In these analyses, the ECAT battery consisted of a test of sequential memory, four tests of spatial ability (including AO), two tests of hand-eye coordination, and a measure of perceptual speed and accuracy.

[11] In the tables that follow, characteristics of items with marked connections will be abbreviated as follows:

Number of connections abbreviated as "No. of Conns"
Number of types of connections as "No. of Types of Conns"
Number of dimensions in figures as "No. of Dimens"
Items containing both 2- and 3-dimensional figures as "2/3"

I. Items With Marked Connections
   A. Shapes of Figures
      - 2-dimensional (simple plane figures)
        - squares, triangles, arrows, hexagons, circles, arches, trapezoids, and parallelograms in approximately equal numbers
      - 3-dimensional (perspective drawings of simple geometric solids)
        - cubes, cylinders, and 3-dimensional triangles, rectangles, and hexagons, in approximately equal numbers
   B. Number of Dimensions in Figures
      - 9 items contain 2-dimensional figures only
      - 1 item contains 3-dimensional figures only
      - 8 items contain both 2- and 3-dimensional figures
   C. Number of Connections
      - 15 items have 3 connections
      - 3 items have 4 connections
   D. Number of Types of Connections (corner, side, inside)
      - 5 items have 1 type
      - 11 items have 2 types
      - 2 items have 3 types

II. Puzzle Items
   A. Puzzle Perimeter Shape (approximately equal numbers)
      - circles
      - squares
      - triangles
      - hexagons
   B. Number of Pieces
      - 2 items have 3 pieces
      - 12 items have 4 pieces
      - 2 items have 5 pieces
      - 2 items have 6 pieces
   C. Does Item Have Shape(s) Embedded In Others?
      - 13 items do not
      - 5 items do

Figure 3. Characteristics of original Assembling Objects items.

I. Items With Marked Connections

	2-DIMENSIONAL			2- AND 3-DIMENSIONAL			
	NUMBER OF TYPES OF CONNECTIONS			NUMBER OF TYPES OF CONNECTIONS			
# CONN	One	Two	Three	One	Two	Three	TOTAL
Three	9	19	2	9	20	1	60
Four	2	3	1	2	4	0	12
TOTAL	11	22	3	11	24	1	72

Note. Specifications for new items followed old item characteristics very closely, except: a) since we thought they might be too difficult, we decided to write fewer items with three types of connections, and b) on similar grounds, we also decided to write no new items with only 3-dimensional figures. Items were written so that mirror images did not affect the correctness of responses.

II. Puzzle Items

	NUMBER OF PIECES				
EMBEDDED	Three	Four	Five	Six	TOTAL
No	6	34	6	6	52
Yes	2	14	2	2	20
TOTAL	8	48	8	8	72

Note. Specifications for new items followed old item characteristics very closely.

Figure 4. Specifications for new Assembling Objects items.

Generation and First Review of Original Item Pool

Working with the final version of the item specifications, the first three authors generated the original pool of 144 draft items, 72 with marked connections and 72 puzzle. These were reviewed by the fourth author and appropriate revisions were made. This first review was generally concerned with the following:

- Making certain that correct answers were indeed correct and that distractors were incorrect
- Making certain that all marked connections were complete (i.e., each connection was shown in the stem on both of the pieces to be connected)
- Discovering and clarifying any ambiguities in the figures (equal vs unequal line lengths, squares vs rectangles, equilateral vs isosceles triangles, etc.)
- Discovering and altering distractors with "incorrect" connections that were very similar to correct ones, which might introduce the unwanted confound of visual acuity into scores
- Discovering and eliminating "mirror images" in items with marked connections (see discussion below)

Issues and Decisions Made

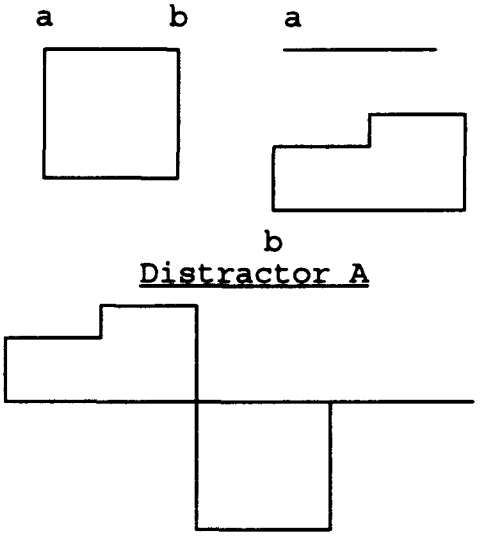
During the process of creating and revising first draft items, we were guided by the item specifications (Figure 4) and a number of other considerations. The issues and decisions involved are discussed separately for each item type below.

Items with Marked Connections. On rational grounds we assumed that the difficulty of items would be directly related to the number of connections among the pieces to be assembled. In addition to following the item specifications for total number of connections, we wanted to create an item pool with approximately the same average number of correct connections per distractor as the items on the original form. An analysis of the original items showed that this average was 1.5 for the items with three marked connections (i.e., a, b, & c) and 2.0 for the items with four (a, b, c, & d). For the 72 draft items, the average number of correct connections per distractor was 1.52 for the 3-connection items and 2.18 for the 4-connection items.[12]

Another issue that came up during the discussions is illustrated with a simple example below. Recall that the

[12]As part of the first review and revision process, two researchers independently counted the number of correct connections in the distractors and correct answers. The level of agreement between the two judges was very high: out of 288 comparisons (4 response alternatives for each of the 72 items), the two judges agreed on 274, or 95.1%. The very few disagreements were then resolved through discussion.

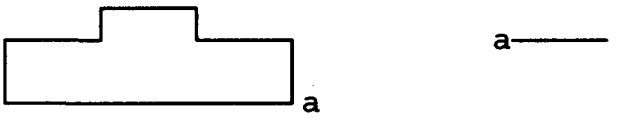
examinee must determine what figure would result if the marked connections were made:



Distractor A

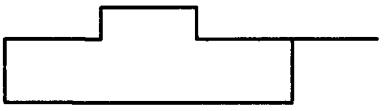
Is the "a" connection correct? The answer is "yes" if we simply note that the line is connected to the square at one corner. The answer is "no" if we consider the relative position of the two connections to the square. [Note that this issue is irrelevant to the "b" connection.] After discussions, we concluded that neither answer is more correct from the standpoint of spatial ability. We therefore decided the issue by a coin toss: connection "a" and all similar ones were to be counted as correct.

We were also concerned about "mirror images" in the items with marked connections. A simple example of this problem is shown below. Once again, the examinee is required to determine what figure would result if the marked connection were made:

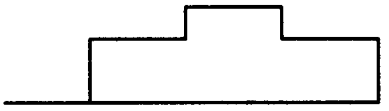


Correct Answer

Distractor A



Distractor B



Distractor A is clearly incorrect; the connection is unambiguously in the wrong location. Distractor B, however, is a

mirror image of the correct response. That is, it would be correct if it where held up to a mirror, or "flipped over" on the page. Since it is known that there are sex differences in examinees' ability to mentally perform three-dimensional rotations of this type (e.g., Alderton, 1989a), we decided that:

- There would be no mirror images in correct responses (that is, such manipulations are not "correct"), and
- No distractors would be incorrect only because they contain a mirror image.

Puzzle Items. An analysis of the original items showed that 28% (5 of 18) have at least one distractor with the wrong number of pieces and that 5% (1 of 18) have more than one distractor with the wrong number of pieces. We followed the first of these parameters - that is, 18 of 72 (25%) of the new items have a distractor with the wrong number of pieces. However, because we believed that such items might be too easy, no new items had more than one distractor with the wrong number of pieces.

First Administration of New Items

Procedure

Researchers created a total of 15 test booklets out of the 144 new items and the 36 original items.[13] Each of the booklets contained 48 items. Twelve of the test booklets contained all 36 original items and a unique subset of 12 new items. Data from these booklets were mainly used to assess the relationship between each new item and the total of the original items. In these booklets, original and new items were interspersed, in groups of six.

To provide further data on the psychometric properties of the new items, an additional three booklets were created, each containing a unique subset of 48 new items. Test instructions for all booklets were identical to the original AO test, as cited above, except that examinees had 52 minutes to complete 48 items.[14]

Sample

Data were collected from 1,549 receptees (i.e., new Army recruits who were about to enter basic training) at Fort Leonard Wood in late 1991. Table 4 describes the assignment of items to

[13]All test booklets also included draft items for another Project A measure. These were placed after the AO items and were timed separately.

[14]When "trying out" items for a power test, test developers commonly take steps to insure that all respondents have ample time to complete all items (Peterson, 1993).

the fifteen booklets, along with the numbers of individuals responding to each booklet. Table 5 shows the total sample sizes for all items.

Table 4

First Administration of New Items: Items and Sample Sizes for Test Booklets

Booklet	Original Items	New Items	N
1	All (001 - 036)	037 - 048	109
2	All	049 - 060	109
3	All	061 - 072	111
4	All	073 - 084	104
5	All	085 - 096	100
6	All	097 - 108	107
7	All	109 - 120	102
8	All	121 - 132	100
9	All	133 - 144	103
10	All	145 - 156	108
11	All	157 - 168	100
12	All	169 - 180	100
13	None	037 - 084	106
14	None	085 - 132	99
15	None	133 - 180	91

Note. Item numbers are those used on original data files, not necessarily those used on the test booklets.

Table 5

First Administration of New Items: Total Sample Sizes for Items

Items		N
Original Items:	001 - 036	1253
New Items:	037 - 048	215
	049 - 060	215
	061 - 072	217
	073 - 084	210
	085 - 096	199
	097 - 108	206
	109 - 120	201
	121 - 132	199
	133 - 144	194
	145 - 156	199
	157 - 168	191
	169 - 180	191

Note. Item numbers are those used on original data files, not necessarily those used on the test booklets.

Results

Tables 6 and 7 show item analysis results for the 72 draft items with marked connections and the 72 draft puzzle items, respectively. Table 8 shows these data for the 36 original items. In all three tables, items are listed by specifications and sorted in terms of highest-to-lowest D_1 value (see below). The following item statistics are shown:

p a measure of item difficulty - the proportion of examinees getting the item correct;

D_1 a measure of item discrimination - the r_{pbis} between the item score (correct or incorrect) and the total score on the 36 original items;

D_2 another measure of item discrimination - the r_{pbis} between the item score and the total score on the 18 original items of the same type (marked connections or puzzles).[15]

Creation of Alternate Forms

Item specifications and item analysis results from the first data collection were used to create three draft forms of the AO test, each composed of 36 new items. In creating these forms, we were guided by the following considerations:

- For each new form, items were chosen so that the average p values [from the first administration of the new items] would be approximately equal to those on the original form.
- Item specifications (Figure 3) were followed so that, subject to other considerations, each new form would be composed of approximately the same number of each type of item (e.g., number of marked connections, number of pieces in puzzles).
- In general, items with the highest D_1 values were chosen. Table 9 shows a summary of item analysis statistics for the items chosen for the three new draft forms.[16]

---

[15]Readers should recall that these item statistics are from individuals who have already been selected into the Army, on the basis of AFQT score and other criteria. One could therefore speculate about how these statistics would differ if this preselection were removed. For example, if AO item scores and total scores are related to AFQT score, then p values may be somewhat lower, and D values higher (due to greater variance in both item and total scores) in an unselected sample.

[16]Appendixes A to C show complete item specification information, as well as psychometric data [from the first administration] for the new items used to construct the new forms.

Table 6

First Administration of New Items: Item Analysis Results, by Specification, for New Items With Marked Connections

No.	Booklet & Item No.	No. of Conns	No. of Types of Conns	No. of Dimens	p	D_1	D_2
106	14 - 34	3	1	2	.76	.55	.56
107	14 - 35	3	1	2	.67	.54	.56
104	14 - 32	3	1	2	.76	.52	.53
055	13 - 7	3	1	2	.84	.51	.56
108	14 - 36	3	1	2	.68	.48	.49
090	14 - 30	3	1	2	.72	.43	.40
038	13 - 2	3	1	2	.86	.42	.46
105	14 - 33	3	1	2	.69	.40	.36
103	14 - 31	3	1	2	.68	.26	.21
056	13 - 8	3	2	2	.73	.56	.54
081	13 - 21	3	2	2	.80	.51	.51
079	13 - 19	3	2	2	.77	.50	.49
042	13 - 6	3	2	2	.79	.45	.53
080	13 - 20	3	2	2	.77	.45	.50
058	13 - 10	3	2	2	.78	.45	.46
037	13 - 1	3	2	2	.83	.45	.39
087	14 - 27	3	2	2	.83	.44	.48
083	13 - 23	3	2	2	.74	.43	.39
065	13 - 17	3	2	2	.81	.30	.27
041	13 - 5	3	2	2	.81	.27	.21
057	13 - 9	3	2	2	.63	.19	.16
085	14 - 25	3	2	2	.05	-.34	-.36
039	13 - 3	3	3	2	.81	.52	.57
063	13 - 15	3	3	2	.88	.52	.44
066	13 - 18	3	3	2	.66	.50	.47
086	14 - 26	3	3	2	.74	.49	.50
082	13 - 22	3	3	2	.67	.40	.35
062	13 - 14	3	3	2	.76	.38	.36
059	13 - 11	3	3	2	.67	.37	.38
088	14 - 28	4	1	2	.71	.51	.56
089	14 - 29	4	1	2	.54	.34	.39
061	13 - 13	4	2	2	.69	.42	.35
064	13 - 16	4	2	2	.74	.29	.27
060	13 - 12	4	3	2	.75	.57	.64
084	13 - 24	4	3	2	.75	.48	.48
040	13 - 4	4	3	2	.61	.45	.45

(Continued)

Table 6 (Continued)

First Administration of New Items: Item Analysis Results, by Specification, for New Items With Marked Connections

No.	Booklet & Item No.	No. of Conns	No. of Types of Conns	No. of Dimens	p	D_1	D_2
112	14 - 40	3	1	2/3	.66	.60	.56
113	14 - 41	3	1	2/3	.52	.53	.54
130	14 - 46	3	1	2/3	.69	.45	.46
151	15 - 7	3	1	2/3	.66	.42	.39
135	15 - 3	3	1	2/3	.54	.34	.36
159	15 - 15	3	1	2/3	.47	.32	.28
133	15 - 1	3	1	2/3	.47	.27	.31
155	15 - 11	3	1	2/3	.71	.16	.15
114	14 - 42	3	2	2/3	.78	.65	.58
132	14 - 48	3	2	2/3	.73	.63	.65
154	15 - 10	3	2	2/3	.78	.60	.56
127	14 - 43	3	2	2/3	.67	.56	.55
129	14 - 45	3	2	2/3	.68	.56	.54
109	14 - 37	3	2	2/3	.82	.56	.51
161	15 - 17	3	2	2/3	.61	.52	.48
158	15 - 14	3	2	2/3	.78	.50	.53
153	15 - 9	3	2	2/3	.79	.49	.51
157	15 - 13	3	2	2/3	.75	.49	.48
136	15 - 4	3	2	2/3	.75	.48	.46
152	15 - 8	3	2	2/3	.71	.43	.46
138	15 - 6	3	2	2/3	.81	.43	.41
110	14 - 38	3	2	2/3	.63	.41	.40
128	14 - 44	3	2	2/3	.66	.40	.38
162	15 - 18	3	2	2/3	.65	.40	.34
160	15 - 16	3	2	2/3	.77	.38	.33
156	15 - 12	3	2	2/3	.71	.35	.34
134	15 - 2	3	2	2/3	.53	.32	.34
137	15 - 5	3	2	2/3	.70	.31	.34
131	14 - 47	3	2	2/3	.26	.12	.14
111	14 - 39	3	3	2/3	.53	.41	.41
176	15 - 20	4	1	2/3	.64	.48	.49
175	15 - 19	4	1	2/3	.66	.42	.48
177	15 - 21	4	2	2/3	.63	.49	.53
178	15 - 22	4	2	2/3	.65	.47	.57
179	15 - 23	4	3	2/3	.60	.45	.49
180	15 - 24	4	3	2/3	.48	.24	.24

Note. In Tables 6, 7, and 8, 'No.' is the item number used on the original data set.

Table 7

First Administration of New Items: Item Analysis Results, by Specification, for New Puzzle Items

No.	Booklet & Item No.	No. of Pieces	Embedded?	p	D_1	D_2
123	14 - 21	3	No	.87	.56	.58
070	13 - 40	3	No	.81	.45	.48
145	15 - 31	3	No	.85	.39	.34
049	13 - 31	3	No	.93	.36	.43
119	14 - 17	3	No	.87	.36	.35
148	15 - 34	3	No	.94	.34	.34
076	13 - 46	3	No	.88	.32	.35
099	14 - 9	3	No	.47	-.00	.10
092	14 - 2	4	No	.75	.63	.62
043	13 - 25	4	No	.90	.58	.55
118	14 - 16	4	No	.89	.55	.56
073	13 - 43	4	No	.92	.53	.49
141	15 - 27	4	No	.84	.51	.47
096	14 - 6	4	No	.81	.50	.51
124	14 - 22	4	No	.85	.50	.51
094	14 - 4	4	No	.83	.49	.55
075	13 - 45	4	No	.92	.47	.43
068	13 - 38	4	No	.58	.46	.47
144	15 - 30	4	No	.77	.46	.46
054	13 - 36	4	No	.93	.44	.43
071	13 - 41	4	No	.72	.44	.41
044	13 - 26	4	No	.86	.43	.43
098	14 - 8	4	No	.92	.43	.36
171	15 - 45	4	No	.91	.43	.31
078	13 - 48	4	No	.88	.42	.36
072	13 - 42	4	No	.70	.41	.44
167	15 - 41	4	No	.86	.39	.39
126	14 - 24	4	No	.72	.37	.35
147	15 - 33	4	No	.57	.37	.29
168	15 - 42	4	No	.85	.36	.41
149	15 - 35	4	No	.92	.35	.47
165	15 - 39	4	No	.90	.35	.41
166	15 - 40	4	No	.83	.33	.44
052	13 - 34	4	No	.93	.33	.31
050	13 - 32	4	No	.93	.27	.36
048	13 - 30	4	No	.77	.27	.32
146	15 - 32	4	No	.91	.26	.31
173	15 - 47	4	No	.93	.26	.29
100	14 - 10	4	No	.94	.23	.19
172	15 - 46	4	No	.91	.21	.14
170	15 - 44	4	No	.87	.21	.11
117	14 - 15	4	No	.72	.18	.18

(Continued)

Table 7 (Continued)

First Administration of New Items: Item Analysis Results, by Specification, for New Puzzle Items

No.	Booklet & Item No.	No. of Pieces	Embedded?	p	D_1	D_2
053	13 - 35	5	No	.85	.52	.56
116	14 - 14	5	No	.76	.47	.50
097	14 - 7	5	No	.89	.46	.40
139	15 - 25	5	No	.90	.32	.35
120	14 - 18	6	No	.83	.48	.50
045	13 - 27	6	No	.83	.44	.44
047	13 - 29	4	Yes	.89	.62	.62
142	15 - 28	4	Yes	.90	.56	.55
164	15 - 38	4	Yes	.86	.55	.53
115	14 - 13	4	Yes	.84	.52	.50
163	15 - 37	4	Yes	.86	.44	.36
121	14 - 19	4	Yes	.84	.41	.38
169	15 - 43	4	Yes	.88	.39	.29
077	13 - 47	4	Yes	.79	.37	.29
051	13 - 33	4	Yes	.93	.35	.41
174	15 - 48	4	Yes	.80	.32	.14
093	14 - 3	4	Yes	.69	.31	.38
069	13 - 39	4	Yes	.82	.31	.36
102	14 - 12	4	Yes	.92	.28	.24
101	14 - 11	4	Yes	.91	.26	.20
074	13 - 44	5	Yes	.87	.66	.67
140	15 - 26	5	Yes	.82	.59	.59
046	13 - 28	5	Yes	.58	.48	.53
125	14 - 23	5	Yes	.84	.43	.43
143	15 - 29	5(6)	Yes	.45	.04	.03
095	14 - 5	6	Yes	.86	.51	.61
091	14 - 1	6	Yes	.87	.48	.54
067	13 - 37	6	Yes	.77	.41	.45
122	14 - 20	6	Yes	.73	.38	.27
150	15 - 36	6	Yes	.69	.30	.38

Note. Item 143 was incorrectly drawn to show 5 pieces, instead of 6.

Table 8

First Administration of New Items: Item Analysis Results, by Specification, for Original Items

Items With Marked Connections

No.	No. of Conns	No. of Types of Conns	No. of Dimens	p	D_1[a]	D_2[a]	D_1	D_2
14	3	1	2	.75	.50	.50	.51	.50
11	3	1	2	.74	.44	.44	.44	.44
12	3	1	2	.62	.43	.42	.43	.42
18	3	2	2	.74	.48	.46	.48	.46
5	3	2	2	.80	.45	.48	.45	.48
3	3	2	2	.81	.41	.43	.41	.44
1	3	2	2	.78	.37	.38	.37	.39
16	4	2	2	.65	.41	.42	.41	.43
13	4	2	2	.48	.25	.24	.25	.24
15	4	2	2	.45	.23	.23	.23	.23
9	3	1	2/3	.85	.48	.50	.48	.52
2	3	1	2/3	.81	.38	.39	.38	.39
8	3	2	2/3	.79	.47	.48	.48	.50
17	3	2	2/3	.71	.43	.42	.43	.42
7	3	2	2/3	.82	.39	.41	.39	.41
10	3	3	2/3	.74	.43	.42	.43	.42
4	4	3	2/3	.15	-.19	-.18	---	---
6	3	2	3	.82	.45	.47	.45	.47

Puzzle Items

No.	No. of Pieces	Embedded?	p	D_1[a]	D_1	D_2
23	3	No	.85	.33	.33	.37
30	4	No	.89	.46	.47	.50
29	4	No	.81	.44	.44	.45
33	4	No	.87	.44	.44	.45
26	4	No	.84	.42	.42	.47
34	4	No	.79	.40	.40	.44
31	4	No	.83	.38	.39	.42
21	4	No	.85	.39	.39	.39
25	4	No	.84	.38	.38	.44
19	4	No	.93	.38	.38	.37
35	4	No	.66	.36	.36	.38
20	4	No	.91	.35	.36	.36
32	5	No	.75	.45	.45	.44
22	5	No	.86	.41	.41	.41
28	6	No	.79	.42	.42	.42
27	3	Yes	.81	.44	.44	.39
24	5	Yes	.35	.20	.21	.19
36	6	Yes	.44	.27	.27	.23

[a]Item 004 was incorrectly drawn and had no correct answer. These statistics are for total scores with this item included; all other statistics are for total scores with the item omitted.

Table 9

First Administration of New Items: Summary of Item Analysis Results for Items Chosen for New Forms

	Form								
	B			C			D		
Statistic	p	D_1	D_2	p	D_1	D_2	p	D_1	D_2
Items With Marked Connections:									
Highest	.86	.57	.64	.88	.60	.57	.84	.65	.65
Lowest	.53	.30	.27	.52	.34	.35	.61	.41	.39
Central	.72	.45	.49	.70	.48	.48	.73	.51	.52
Puzzle Items:									
Highest	.93	.62	.62	.93	.66	.67	.94	.59	.59
Lowest	.57	.31	.29	.58	.35	.31	.58	.30	.34
Central	.84	.44	.43	.83	.45	.46	.83	.45	.45

Note. Central values of p are averages; central values of D are medians. Overall average p values are: Form B - .77, Form C - .74, Form D - .76. Average p values for the original items were: Marked Connections - .73, Puzzles - .83, Overall - .78.

Second Administration of New Items

A second data collection was conducted to assess the psychometric properties of the three draft forms and their statistical relationship with the original form.

Procedure

Each examinee took the original form and one of the three draft forms. To control for order effects, approximately half the subjects received the original form first and the draft form second; the other subjects received the draft form first, followed by the original form. The forms were administered from separate test booklets. Test instructions for all booklets were identical to the original AO test, as cited above, except that examinees had 40 minutes to complete the 36 items on each form.

Sample

Data were collected from 2,520 receptees at Fort Leonard Wood between February and April, 1992.

Results

Order Effects. T tests and Chi-square tests were run to determine any significant order effects upon total scores and item scores, respectively, for the original form (A) and the new forms (B,C,D). As Table 10 shows, very few of these effects were significant; in all cases, the magnitude of significant effects was rather small. Because of this, we collapsed the data, across order, for all subsequent analyses.

Table 10

Second Administration of New Items: Tests of Order Effects

Sample	Type of Score	Form	Significant Order Effects
A - B	Total	A	None
	Total	B	None
	Item	A	Nos. 7, 12, 14, 30
	Item	B	None
A - C	Total	A	None
	Total	C	None
	Item	A	None
	Item	C	None
A - D	Total	A	None
	Total	D	Total, Marked
	Item	A	Nos. 3, 26, 30
	Item	D	Nos. 1, 3, 11, 17, 32

Note. T-tests were used to test total score effects; Chi-square tests were employed to test item score effects. All results are for $p<.05$

Item Analyses. Tables 11 to 13 show item analysis results (p, D_1, and D_2 values, as defined above) for the second administration of the new items, grouped into Forms B, C, and D, respectively. As can be seen, p values were all within a reasonable range (.45 to .92) and D_1 and D_2 values for all items approached or exceeded .3. The results in Tables 11 to 13 are summarized in Table 14.

Table 15 shows two types of part-whole (item-total) correlations for the new items and forms: "Total" refers to the total score for all items on the form; "Marked" and "Puzzle" refer to the total scores on items with marked connections and puzzle items, respectively. As shown, all indices approached or

Table 11

Second Administration of New Items: Item Analysis Results for Form B

Items With Marked Connections

Item No.	Old Booklet & Item No.	No. of Conns	No. of Types of Conns	No. of Dimens	p	D_1	D_2
1	13 - 02	3	1	2	.75	.33	.34
2	13 - 17	3	2	2	.81	.29	.29
3	13 - 06	3	2	2	.78	.40	.42
4	13 - 20	3	2	2	.82	.29	.27
5	13 - 03	3	3	2	.81	.39	.40
6	14 - 34	3	1	2	.78	.36	.36
7	13 - 12	4	3	2	.70	.40	.38
8	14 - 46	3	1	2/3	.70	.44	.43
9	14 - 43	3	2	2/3	.74	.38	.38
10	15 - 23	4	3	2/3	.71	.33	.31
11	14 - 45	3	2	2/3	.71	.37	.36
12	15 - 07	3	1	2/3	.65	.44	.39
13	14 - 30	3	1	2	.73	.35	.34
14	15 - 04	3	2	2/3	.71	.34	.31
15	15 - 18	3	2	2/3	.71	.35	.34
16	14 - 39	3	3	2/3	.54	.32	.26
17	15 - 13	3	2	2/3	.70	.40	.39
18	13 - 14	3	3	2	.74	.44	.41

Puzzle Items

Item No.	Old Booklet & Item No.	No. of Pieces	Embedded?	p	D_1	D_2
19	13 - 31	3	No	.89	.33	.37
20	14 - 08	4	No	.88	.31	.35
21	13 - 45	4	No	.90	.34	.38
22	13 - 25	4	No	.91	.37	.39
23	14 - 16	4	No	.92	.39	.41
24	13 - 39	4	Yes	.84	.34	.35
25	13 - 26	4	No	.87	.42	.43
26	13 - 35	5	No	.82	.43	.44
27	14 - 01	6	Yes	.84	.39	.42
28	13 - 29	4	Yes	.87	.38	.38
29	14 - 05	6	Yes	.80	.44	.45
30	14 - 04	4	No	.84	.37	.41
31	15 - 31	3	No	.81	.37	.37
32	14 - 24	4	No	.64	.30	.33
33	13 - 27	6	No	.74	.35	.37
34	14 - 03	4	Yes	.68	.33	.33
35	14 - 19	4	Yes	.80	.31	.32
36	15 - 33	4	No	.63	.28	.27

Table 12

Second Administration of New Items: Item Analysis Results for Form C

Items With Marked Connections

Item No.	Old Booklet & Item No.	No. of Conns	No. of Types of Conns	No. of Dimens	p	D_1	D_2
1	13 - 15	3	3	2	.85	.41	.44
2	13 - 10	3	2	2	.76	.41	.42
3	13 - 19	3	2	2	.72	.34	.32
4	14 - 37	3	2	2/3	.77	.30	.31
5	15 - 09	3	2	2/3	.80	.40	.39
6	13 - 13	4	2	2	.72	.41	.41
7	15 - 03	3	1	2/3	.45	.29	.29
8	15 - 22	4	2	2/3	.73	.37	.34
9	14 - 36	3	1	2	.70	.46	.47
10	14 - 35	3	1	2	.71	.50	.48
11	13 - 22	3	3	2	.58	.36	.36
12	13 - 23	3	2	2	.67	.36	.33
13	14 - 44	3	2	2/3	.74	.40	.39
14	14 - 33	3	1	2	.61	.33	.30
15	15 - 10	3	2	2/3	.76	.44	.42
16	15 - 20	4	1	2/3	.55	.43	.40
17	15 - 08	3	2	2/3	.63	.37	.33
18	14 - 41	3	1	2/3	.60	.38	.36

Puzzle Items

Item No.	Old Booklet & Item No.	No. of Pieces	Embedded?	p	D_1	D_2
19	14 - 17	3	No	.87	.33	.34
20	13 - 43	4	No	.88	.46	.44
21	15 - 45	4	No	.89	.43	.45
22	13 - 33	4	Yes	.83	.36	.35
23	15 - 28	4	Yes	.89	.41	.40
24	14 - 22	4	No	.83	.35	.34
25	15 - 41	4	No	.82	.38	.38
26	14 - 02	4	No	.78	.43	.39
27	14 - 23	5	Yes	.78	.42	.41
28	13 - 37	6	Yes	.79	.40	.38
29	14 - 18	6	No	.80	.40	.38
30	13 - 44	5	Yes	.77	.46	.42
31	13 - 40	3	No	.77	.40	.37
32	15 - 42	4	No	.82	.43	.41
33	14 - 14	5	No	.70	.40	.40
34	13 - 41	4	No	.67	.33	.30
35	13 - 38	4	No	.47	.29	.27
36	15 - 37	4	Yes	.76	.41	.40

Table 13

Second Administration of New Items: Item Analysis Results for Form D

Items With Marked Connections

Item No.	Old Booklet & Item No.	No. of Conns	No. of Types of Conns	No. of Dimens	p	D_1	D_2
1	13 - 07	3	1	2	.84	.39	.43
2	13 - 01	3	2	2	.83	.37	.37
3	14 - 27	3	2	2	.80	.39	.42
4	13 - 21	3	2	2	.81	.33	.34
5	15 - 06	3	2	2/3	.81	.43	.40
6	14 - 42	3	2	2/3	.83	.45	.43
7	15 - 19	4	1	2/3	.69	.31	.31
8	14 - 32	3	1	2	.74	.44	.43
9	15 - 21	4	2	2/3	.68	.34	.34
10	14 - 38	3	2	2/3	.64	.37	.35
11	15 - 17	3	2	2/3	.52	.31	.31
12	14 - 26	3	3	2	.74	.35	.33
13	14 - 40	3	1	2/3	.64	.37	.36
14	14 - 28	4	1	2	.67	.39	.37
15	13 - 18	3	3	2	.65	.38	.37
16	15 - 14	3	2	2/3	.67	.41	.39
17	14 - 48	3	2	2/3	.73	.40	.37
18	13 - 08	3	2	2	.72	.41	.38

Puzzle Items

Item No.	Old Booklet & Item No.	No. of Pieces	Embedded?	p	D_1	D_2
19	15 - 34	3	No	.91	.32	.34
20	13 - 46	3	No	.87	.34	.36
21	15 - 35	4	No	.91	.32	.32
22	13 - 36	4	No	.91	.37	.36
23	15 - 39	4	No	.91	.30	.29
24	14 - 13	4	Yes	.85	.38	.37
25	14 - 21	3	No	.81	.32	.32
26	14 - 06	4	No	.86	.39	.41
27	13 - 28	5	Yes	.55	.27	.25
28	15 - 30	4	No	.79	.30	.30
29	13 - 42	4	No	.52	.33	.33
30	15 - 36	6	Yes	.57	.32	.31
31	15 - 38	4	Yes	.87	.34	.35
32	14 - 07	5	No	.78	.43	.46
33	15 - 27	4	No	.77	.45	.45
34	15 - 26	5	Yes	.77	.43	.42
35	13 - 48	4	No	.75	.45	.45
36	15 - 25	5	No	.77	.34	.36

Table 14

Second Administration of New Items: Summary of Item Analysis Results

	Form								
	B			C			D		
Statistic	p	D_1	D_2	p	D_1	D_2	p	D_1	D_2
Items With Marked Connections:									
Highest	.82	.44	.43	.85	.50	.48	.84	.45	.43
Lowest	.54	.29	.26	.45	.29	.29	.52	.31	.31
Central	.73	.37	.36	.69	.39	.38	.72	.39	.37
Puzzle Items:									
Highest	.92	.44	.45	.89	.46	.45	.91	.45	.46
Lowest	.63	.28	.27	.47	.29	.27	.52	.27	.25
Central	.82	.36	.38	.79	.40	.39	.79	.34	.36

Note. Central values of p are averages; central values of D are medians. Overall average p values are: Form B - .77, Form C - .74, Form D - .76.

Table 15

Second Administration of New Items: Part-Whole (Item-total) Correlations for New Forms

Items With Marked Connections:

Item	Form B		Form C		Form D	
No.	Total	Marked	Total	Marked	Total	Marked
01	.353	.423	.437	.455	.485	.540
02	.325	.387	.423	.401	.419	.461
03	.452	.514	.378	.392	.475	.528
04	.336	.345	.297	.350	.434	.473
05	.394	.441	.426	.463	.495	.507
06	.454	.447	.442	.469	.467	.485
07	.439	.492	.305	.306	.348	.349
08	.499	.521	.423	.438	.465	.481
09	.420	.451	.482	.491	.365	.380
10	.395	.415	.547	.534	.403	.447
11	.448	.465	.366	.404	.304	.347
12	.430	.413	.381	.387	.455	.475
13	.423	.458	.437	.446	.459	.471
14	.386	.417	.364	.361	.474	.474
15	.415	.424	.469	.478	.442	.444
16	.349	.345	.445	.430	.479	.488
17	.501	.481	.390	.393	.490	.477
18	.542	.520	.406	.388	.498	.470

Puzzle Items:

Item	Form B		Form C		Form D	
No.	Total	Puzzle	Total	Puzzle	Total	Puzzle
19	.463	.508	.407	.440	.417	.418
20	.441	.498	.563	.558	.457	.507
21	.479	.541	.543	.563	.463	.532
22	.515	.569	.435	.465	.512	.571
23	.527	.537	.507	.558	.486	.532
24	.435	.475	.427	.480	.498	.528
25	.527	.593	.478	.506	.429	.468
26	.523	.549	.518	.538	.531	.587
27	.514	.553	.508	.560	.310	.313
28	.552	.602	.460	.480	.396	.433
29	.543	.566	.535	.567	.338	.307
30	.460	.537	.564	.570	.346	.339
31	.454	.484	.480	.522	.471	.495
32	.369	.405	.507	.531	.518	.548
33	.444	.474	.477	.495	.541	.585
34	.413	.425	.379	.377	.511	.555
35	.435	.457	.327	.316	.521	.559
36	.351	.392	.523	.551	.430	.468

exceeded .3 and many were greater than .5.[17] The results in Table 15 are summarized in Table 16.

Table 16

Second Administration of New Items: Summary of Part-Whole (Item-total) Correlations

	Form					
	B		C		D	
	Total	Type	Total	Type	Total	Type
Items With Marked Connections:						
Highest	.542	.521	.547	.534	.498	.540
Lowest	.325	.345	.297	.306	.304	.347
Median	.422	.444	.423	.417	.462	.474
Puzzle Items:						
Highest	.552	.602	.564	.570	.541	.585
Lowest	.351	.392	.327	.316	.310	.307
Median	.462	.523	.494	.527	.467	.518

Note. 'Total' refers to the score across all items; 'Type' refers to the score across items of the same type (marked connections or puzzle).

Descriptive Data and Analyses of Reliability. Table 17 shows descriptive and correlational data for the new and original forms. These data include a number of reliability estimates, as follows:

1) internal consistency, shown as alpha coefficients in the diagonals
2) alternate forms, shown as correlations between scores on Form A and those on the new forms, which are also shown in Table 18.

[17]According to Nunnally (1978), item-total correlations exceeding .30 are generally considered to be good. Another popular rule of thumb, as provided by Crocker and Algina (1986), is to retain items that exceed zero by at least two standard error units. Applying the correct formulas to our very large samples, we calculate that minimal acceptable correlations would range from .068 to .069.

Table 17

Second Administration of New Items: Descriptive Data

Forms A and B (N=835)

Measure	M	SD	Intercorrelation with: (1)	(2)	(3)	(4)	(5)	(6)
(1) Form A (Total)	25.77	6.66	.875					
(2) Form A (Marked)	12.58	3.90	.898	.810				
(3) Form A (Puzzle)	13.19	3.60	.878	.578	.809			
(4) Form B (Total)	27.77	7.12	.745	.645	.680	.905		
(5) Form B (Marked)	13.08	4.16	.700	.675	.564	.897	.845	
(6) Form B (Puzzle)	14.69	3.86	.620	.462	.647	.878	.576	.876

Forms A and C (N=862)

Measure	M	SD	Intercorrelation with: (1)	(2)	(3)	(4)	(5)	(6)
(1) Form A (Total)	26.12	6.70	.881					
(2) Form A (Marked)	12.72	3.96	.901	.822				
(3) Form A (Puzzle)	13.40	3.57	.877	.582	.815			
(4) Form C (Total)	26.51	7.44	.797	.718	.699	.906		
(5) Form C (Marked)	12.37	4.13	.760	.736	.609	.907	.829	
(6) Form C (Puzzle)	14.13	4.08	.684	.565	.657	.905	.642	.876

Forms A and D (N=823)

Measure	M	SD	Intercorrelation with: (1)	(2)	(3)	(4)	(5)	(6)
(1) Form A (Total)	26.21	6.50	.873					
(2) Form A (Marked)	12.75	3.86	.898	.811				
(3) Form A (Puzzle)	13.46	3.47	.872	.567	.805			
(4) Form D (Total)	27.18	7.25	.753	.667	.667	.906		
(5) Form D (Marked)	13.01	4.24	.706	.687	.555	.907	.854	
(6) Form D (Puzzle)	14.17	3.84	.642	.499	.646	.885	.607	.861

Form A - Total Sample (N=2520)

Measure	M	SD	Intercorrelation with: (1)	(2)	(3)
(1) Form A (Total)	26.03	6.62	.876		
(2) Form A (Marked)	12.68	3.91	.899	.815	
(3) Form A (Puzzle)	13.35	3.55	.876	.576	.810

Note. All intercorrelations are significant at p<.0001. Alpha coefficients are shown on the diagonals, and are underlined.

Table 18

Second Administration of New Items: Alternate Forms Reliability Results

	Form		
	B	C	D
Items with Marked Connections	.675 [.816]	.736 [.892]	.687 [.826]
Puzzle Items	.647 [.769]	.657 [.778]	.646 [.776]
Total	.745 [.837]	.797 [.892]	.753 [.847]

Note. Figures in the table are correlations between scores on each new form and the original. Those in brackets are corrected for each form's internal consistency (alpha coefficients are shown in the diagonals in Table 17).

According to a number of sources, these results demonstrate acceptable internal consistency and alternate forms reliability (Nunnally, 1978; Crocker & Algina, 1986).[18]

Discussion

The Project A Assembling Objects is a strong candidate for inclusion in future versions of the ASVAB. In anticipation of future needs, we created specifications for generating new AO items. These specifications were validated when we were able to use them in developing a pool of 108 new items that measure the same construct as the original instrument. We then combined the items into three entire new forms of AO, each possessing good psychometric properties. Thus, the new items can be used in fixed test forms or as a pool to be sampled from in adaptive testing. Additionally, our results should be a good start on future research, development, and analyses, including:

[18]Regarding internal consistency reliability, Nunnally (1978) has advised that alpha coefficients should be at least .7 to .9, depending on the instrument's intended use (e.g., coefficients for selection tests should be higher than those for research instruments).

Referring to alternate forms reliability, Crocker & Algina (1986, p. 132) stated that "Although there are no hard, fast rules for what constitutes a minimally acceptable value..., many standardized achievement test manuals report coefficients ranging in the .80s and .90s for this type of reliability." More specifically, Hartigan and Wigdor (1989) reported alternate form reliabilities for subtests from ASVAB Forms 8, 9, and 10 ranging from .57 to .90 with a median of .79. The same authors reported alternate form reliability estimates for the GATB spatial subtest ranging from .78 to .85 for intertest intervals of up to one week.

- Assessing the impact of various item parameters (from our item specifications and other sources) upon item difficulty and perhaps also discrimination[19]
- Determining any ethnic or sex bias in the new items and forms[20]
- Collecting normative data on suitably large and representative samples
- Equating scores across our new forms and others that may be created
- Developing more items for use in computer adaptive testing, using our validated item specifications
- Creating forms of different lengths suitable for inclusion in future versions of ASVAB.[21]

The Assembling Objects test is an excellent measure of both spatial ability and more general cognitive skills. It should therefore be a worthy addition to the selection and classification systems of the Army and the other armed services. We hope our efforts have contributed to its long term validity and usefulness.

---

[19]Although readers may choose to examine the tables in the present report for some preliminary indications of these relationships, it should be borne in mind that the data shown were not collected under operational, timed conditions.

[20]Campbell and Zook (1992) have reported normative data on the original AO test, collected during the Concurrent and Longitudinal Validations of the Project A measures. Ranges of effect sizes (mean differences divided by the male or white standard deviation) were, Female: -.02 to .08; Black: .82 to .86; Hispanic: .15 to .25; Other: .09 to .15.

[21]Due to certain practical considerations, it is very doubtful that the full, 36-item version of AO will be used in future versions of ASVAB.

REFERENCES

Alderton, D. L. (1989a, March). The fleeting nature of sex differences in spatial ability. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Alderton, D. L. (1989b). Development and evaluation of integrating details: A complex spatial problem solving test (NPRDC Technical Report 89-6). San Diego, CA: Navy Personnel Research and Development Center.

Anastasi, A. (1982). Psychological testing (5th ed.). New York: MacMillan.

Busciglio, H. H. (1990). The incremental validity of spatial and perceptual-psychomotor tests relative to the Armed Services Vocational Aptitude Battery (ARI Technical Report 883). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A220 903)

Busciglio, H. H. (1991). The usefulness of Project A spatial tests for predicting comprehensive performance measures (ARI Research Note 91-17). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A232 069)

Campbell, J. P. (Ed.). (1989). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1987 fiscal year (ARI Technical Report 862). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A219 046)

Campbell, J. P., & Zook, L. M. (Eds.). (1992). Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel: Annual report, 1990 fiscal year (ARI Technical Report 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A252 675)

Carey, N. B. (1992, August). New predictors of mechanics job performance: Marine Corps findings. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston.

Flanagan, J. C. (1965). Flanagan Industrial Test manual. Chicago: Science Research Associates.

Hartigan, J. A., & Wigdor, A. K. (Eds.). (1989). Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery. Washington, DC: National Academy Press.

Likert, R., & Quasha, W. H. (1970). Manual for the Revised Minnesota Paper Form Board Test. New York: The Psychological Corporation.

Lohman, D. F. (1979). Spatial ability: A review and reanalysis of the correlational literature. Technical Report No. 8, Aptitude Research Project, School of Education, Stanford University.

Nunnally, J. C. (1978). Psychometric theory (2nd Ed.). New York: McGraw-Hill.

Peterson, N. G. (Ed.). (1987). Development and field test of the trial battery for Project A (ARI Technical Report 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A184 575)

Peterson, N. G. (1993). Personal communication.

Ruch, F. L., & Ruch, W. W. (1980). Employee Aptitude Survey: Technical report. Los Angeles, CA: Psychological Services, Inc.

APPENDIX A

Characteristics of Items Chosen for Form B

Items With Marked Connections

Item No.	Old Booklet & Item No.	No. of Conns	No. of Types of Conns	No. of Dimens	p	D_1	D_2
1	13 - 02	3	1	2	.86	.42	.46
2	13 - 17	3	2	2	.81	.30	.27
3	13 - 06	3	2	2	.79	.45	.53
4	13 - 20	3	2	2	.77	.45	.50
5	13 - 03	3	3	2	.81	.52	.57
6	14 - 34	3	1	2	.76	.55	.56
7	13 - 12	4	3	2	.75	.57	.64
8	14 - 46	3	1	2/3	.69	.45	.46
9	14 - 43	3	2	2/3	.67	.56	.55
10	15 - 23	4	3	2/3	.60	.45	.49
11	14 - 45	3	2	2/3	.68	.56	.54
12	15 - 07	3	1	2/3	.66	.42	.39
13	14 - 30	3	1	2	.72	.43	.40
14	15 - 04	3	2	2/3	.75	.48	.46
15	15 - 18	3	2	2/3	.65	.40	.34
16	14 - 39	3	3	2/3	.53	.41	.41
17	15 - 13	3	2	2/3	.75	.49	.48
18	13 - 14	3	3	2	.76	.38	.36

Puzzle Items

Item No.	Old Booklet & Item No.	No. of Pieces	Embedded?	p	D_1	D_2
19	13 - 31	3	No	.93	.36	.43
20	14 - 08	4	No	.92	.43	.36
21	13 - 45	4	No	.92	.47	.43
22	13 - 25	4	No	.90	.58	.55
23	14 - 16	4	No	.89	.55	.56
24	13 - 39	4	Yes	.82	.31	.36
25	13 - 26	4	No	.86	.43	.43
26	13 - 35	5	No	.85	.52	.56
27	14 - 01	6	Yes	.87	.48	.54
28	13 - 29	4	Yes	.89	.62	.62
29	14 - 05	6	Yes	.86	.51	.61
30	14 - 04	4	No	.83	.49	.55
31	15 - 31	3	No	.85	.39	.34
32	14 - 24	4	No	.72	.37	.35
33	13 - 27	6	No	.83	.44	.44
34	14 - 03	4	Yes	.69	.31	.38
35	14 - 19	4	Yes	.84	.41	.38
36	15 - 33	4	No	.57	.37	.29

APPENDIX B

Characteristics of Items Chosen for Form C

Items With Marked Connections

Item No.	Old Booklet & Item No.	No. of Conns	No. of Types of Conns	No. of Dimens	p	D_1	D_2
1	13 - 15	3	3	2	.88	.52	.44
2	13 - 10	3	2	2	.78	.45	.46
3	13 - 19	3	2	2	.77	.50	.49
4	14 - 37	3	2	2/3	.82	.56	.51
5	15 - 09	3	2	2/3	.79	.49	.51
6	13 - 13	4	2	2	.69	.42	.35
7	15 - 03	3	1	2/3	.54	.34	.36
8	15 - 22	4	2	2/3	.65	.47	.57
9	14 - 36	3	1	2	.68	.48	.49
10	14 - 35	3	1	2	.67	.54	.56
11	13 - 22	3	3	2	.67	.40	.35
12	13 - 23	3	2	2	.74	.43	.39
13	14 - 44	3	2	2/3	.66	.40	.38
14	14 - 33	3	1	2	.69	.40	.36
15	15 - 10	3	2	2/3	.78	.60	.56
16	15 - 20	4	1	2/3	.64	.48	.49
17	15 - 08	3	2	2/3	.71	.43	.46
18	14 - 41	3	1	2/3	.52	.53	.54

Puzzle Items

Item No.	Old Booklet & Item No.	No. of Pieces	Embedded?	p	D_1	D_2
19	14 - 17	3	No	.87	.36	.35
20	13 - 43	4	No	.92	.53	.49
21	15 - 45	4	No	.91	.43	.31
22	13 - 33	4	Yes	.93	.35	.41
23	15 - 28	4	Yes	.90	.56	.55
24	14 - 22	4	No	.85	.50	.51
25	15 - 41	4	No	.86	.39	.39
26	14 - 02	4	No	.75	.63	.62
27	14 - 23	5	Yes	.84	.43	.43
28	13 - 37	6	Yes	.77	.41	.45
29	14 - 18	6	No	.83	.48	.50
30	13 - 44	5	Yes	.87	.66	.67
31	13 - 40	3	No	.81	.45	.48
32	15 - 42	4	No	.85	.36	.41
33	14 - 14	5	No	.76	.47	.50
34	13 - 41	4	No	.72	.44	.41
35	13 - 38	4	No	.58	.46	.47
36	15 - 37	4	Yes	.86	.44	.36

APPENDIX C

Characteristics of Items Chosen for Form D

Items With Marked Connections

Item No.	Old Booklet & Item No.	No. of Conns	No. of Types of Conns	No. of Dimens	p	D_1	D_2
1	13 - 07	3	1	2	.84	.51	.56
2	13 - 01	3	2	2	.83	.45	.39
3	14 - 27	3	2	2	.83	.44	.48
4	13 - 21	3	2	2	.80	.51	.51
5	15 - 06	3	2	2/3	.81	.43	.41
6	14 - 42	3	2	2/3	.78	.65	.58
7	15 - 19	4	1	2/3	.66	.42	.48
8	14 - 32	3	1	2	.76	.52	.53
9	15 - 21	4	2	2/3	.63	.49	.53
10	14 - 38	3	2	2/3	.63	.41	.40
11	15 - 17	3	2	2/3	.61	.52	.48
12	14 - 26	3	3	2	.74	.49	.50
13	14 - 40	3	1	2/3	.66	.60	.56
14	14 - 28	4	1	2	.71	.51	.56
15	13 - 18	3	3	2	.66	.50	.47
16	15 - 14	3	2	2/3	.78	.50	.53
17	14 - 48	3	2	2/3	.73	.63	.65
18	13 - 08	3	2	2	.73	.56	.54

Puzzle Items

Item No.	Old Booklet & Item No.	No. of Pieces	Embedded?	p	D_1	D_2
19	15 - 34	3	No	.94	.34	.34
20	13 - 46	3	No	.88	.32	.35
21	15 - 35	4	No	.92	.35	.47
22	13 - 36	4	No	.93	.44	.43
23	15 - 39	4	No	.90	.35	.41
24	14 - 13	4	Yes	.84	.52	.50
25	14 - 21	3	No	.87	.56	.58
26	14 - 06	4	No	.81	.50	.51
27	13 - 28	5	Yes	.58	.48	.53
28	15 - 30	4	No	.77	.46	.46
29	13 - 42	4	No	.70	.41	.44
30	15 - 36	6	Yes	.69	.30	.38
31	15 - 38	4	Yes	.86	.55	.53
32	14 - 07	5	No	.89	.46	.40
33	15 - 27	4	No	.84	.51	.47
34	15 - 26	5	Yes	.82	.59	.59
35	13 - 48	4	No	.88	.42	.36
36	15 - 25	5	No	.90	.32	.35